DOCUMENT RESUME

ED 300 400 TM 012 377

AUTHOR Emihovich, Catherine; And Others
TITLE The Making of a Model: Utilizing Consensus in Formative Evaluation.
PUB DATE Oct 87
NOTE 14p.; Paper presented at the Annual Meeting of the American Evaluation Association (Boston, MA, October 14-17, 1987).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Case Studies; Evaluation Criteria; *Evaluation Methods; Evaluators; *Formative Evaluation; Instructional Leadership; Performance Factors; Personnel Evaluation; Principals; *Program Evaluation; Recognition (Achievement); State Programs
IDENTIFIERS *Consensus Models; Principal Incentive Program; *Stakeholder Evaluation

ABSTRACT

A case study description of work toward developing an evaluation model for implementation of a statewide principal incentive program illustrates the achieving of consensus when multiple stakeholders' interests were involved. The type of training and background needed by evaluators are also considered. In response to the Educational Improvement Act of 1984 (South Carolina), principal improvement programs were initiated, requiring an evaluation model. The first task was to select criteria and performance indicators. Extensive discussion by a committee of educators and community representatives reduced the list to 25 indicators. The next task was identifying appropriate evidence for judging performance. Use of small groups helped build consensus. The model as constituted was evaluated with a hypothetical case study. The process suggested that successful evaluators need diversified training and an understanding of evaluation in multiple contexts. Models that make sense to stakeholders should be used. The outlined model includes the dimensions of student achievement, leadership, staff supervision, community relations, and personal development. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED300400

The making of a model: Utilizing consensus in formative evaluation

Catherine Eminovich
Florida State University
Department of Educational Research
College of Education
Tallahassee, FL 32306
(904) 644-4592

Michael Rowls
Department of Instruction & Teacher Education
George Lackey
Department of Educational Psychology
University of South Carolina
College of Education
Columbia, SC 29208

Paper presented at the American Evaluation Association Annual Meeting, October 15-17, at Boston, Massachusetts

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

CATHERINE EMINOVICH

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM012 377

One means of ensuring that an evaluation model will be accepted and the results utilized is if the participants or clients have a stake in developing the model itself. Weiss (1986b) noted that this is one of the primary assumptions of the stakeholder approach, in that it assumes that "stakeholders want to have evaluative information about the program and they are willing to participate in the evaluative process" (Weiss, 1986, p. 187). But the issue of stakeholder participation raises several questions which are the focus of this paper: (1) how is consensus achieved when there are multiple stakeholders' interests' involved?; (2) what factors need to be considered in designing an evaluation model which addresses both stakeholders' local concerns and meets the state education department's policy objectives for the statewide program?; and, (3) what type of training and background do evaluators need to make the stakeholder approach successful?

These questions are answered through a case study description of the authors' work in developing an evaluation model for implementation of a statewide principal incentive program. The description is presented from three perspectives: origins of the model; evaluative dimensions of the model; and, future training of evaluators. In the first section, the social dynamics underlying the authors' work with a group comprised of community leaders and educational personnel from a small school district is described as they struggled to achieve consensus on the criteria for evaluating superior principals. This formative evaluation process is also examined in relation to the assumptions of both the stakeholder and responsive evaluation models as to how well these models worked in a real-life setting.

The second section of the paper focuses on how the model could have been implemented in all districts across the state. The authors' view is that to create a successful model which can be used beyond a local context, the model must work on several levels of meaning brought into play by the various constituencies likely to be affected by the model. The utilization of such a model depends upon the use of multiple methods for assessment (both quantitative and qualitative) as well as the evaluator's understanding of the sociopolitical context of evaluation. Patton (1986) noted this latter aspect was one of the most important predictors of whether evaluation decisions would be used by clients.

The third section of the paper suggests that the training of evaluators should include extensive preparation in multiple methods, exposure to ideologies of competing research paradigms, and application of evaluation models to the appropriate contexts. This approach to training would reflect ideas recently expressed in instructional psychology that students be taught how to use strategies for solving problems rather than how to memorize discrete facts. In a similar fashion, evaluators would be taught how to apply their professional knowledge to actual problems in the field, rather than just focusing on learning a single paradigm or set of methods. The paper concludes by considering the question that the mixing of models may in fact become the model for a successful evaluator to follow.

I. Beginnings

In 1984 South Carolina passed the Educational Improvement Act, a

landmark piece of legislation designed to improve the schools and raise the state's prestige in educational leadership. A component of the act called for the identification of superior teachers and principals, who would receive incentive pay and other forms of recognition for their work. In 1985 the State Department of Education worked on identifying criteria for superior teachers, and in 1986 the DOE put out nine RFP's for districts to respond to in developing a model for identifying the superior principal. Each district would receive between $25,000-30,000 to develop a model which would be submitted to a review board, who would then select two or three of the models for state-wide implementation the following year.

The districts who chose to apply for this money received very few guidelines from DOE as to how the models should be constructed, although the DOE did maintain a gatekeeping role by stipulating that before principals could apply for incentive money, they must first meet the following conditions: (1) receive a superior score on the DOE's own evaluation instrument; (2) their school must have demonstrated either a mean averaged gain on both BSAP (a test of basic skills developed in-state under the Basic Skills Assessment Program) and CTBS (a nationally norm-referenced test) with a significant z-score change, or a maintenance of a positive achievement trend from the previous year; and, (3) the principals must choose to apply voluntarily. With these conditions, the DOE remained very much a stakeholder in the overall process even though the officials insisted they wanted the districts to develop models which were tailored to local operating conditions.

One of the districts who chose to apply was a small district down in the southern part of the state near the city of Charleston. The Assistant Superintendent of Instruction contacted the third author, a professor of educational research at the University of South Carolina, and asked him to help her submit a proposal to secure funds to develop a principal incentive model for her district. Their application was successful, and early in December 1985 the first meeting of the Model Development Committee (MDC) was convened in Hilton Head, SC. The 24 committee members (not all of whom were present at this first meeting) consisted of people from academia (five professors and two graduate assistants), the school district (the district superintendent, two assistants superintendents, four principals, three teachers and three board members), and four community representatives. The MDC was deliberately selected to obtain a broad cross-section of views for building an effective and parsimonious model based on the experiences and ideas of a talented group of people interested in the district's educational progress.

When the MDC first met, the general perception among the school and community members was that this would be another "academic" exercise, "all talk and no action." They were pleasantly surprised to be handed an agenda with specific objectives to be accomplished during this first meeting. These objectives were:

> (1) to orient committee members to the nature of the principal incentive project and strategies to be employed in meeting the objectives of the project;

(2) to identify a pool of acceptable and desirable incentives to be used to reward superior principals;

(3) to identify general criteria and characteristics for evaluating the professional performance of principals; and,

(4) to identify community and school groups (i.e., constituency review groups) who will review the work of the Model Development Committee with respect to the specific elements of principal incentives and evaluation performance criteria

Another strategy which was successfully employed to help build consensus was to divide the committee into small groups, each of which consisted of a teacher, administrator, board member, academic, and community member, to work on their assignments. Early in the meeting, the third author observed that the administrators and board members dominated the discussion, and that the classroom teachers (all of whom were women; only one board member was also a woman) were reluctant to give their opinions. By ensuring that the teachers were given a forum in a more congenial setting (e.g., a small group), this led to their increased visibility in the general meetings. By the end of the first two day conference, the group had gained a cohesive sense of identity and purpose and left with the feeling that this model development would be a productive experience. As one of the professors noted, the committee would be engaged in developing "successive approximations" until a final model was produced.

II. The Making of a Model

The actual construction of the model took place during the second conference and third conferences. The second conference was held during the second week of January, 1986, and the major objectives for this conference were:

(1) to report the results of the survey conducted with 215 district personnel and community members regarding appropriate incentives and performance indicators;

(2) to describe procedures used in analyzing and synthesizing the information obtained from the total constituency survey;

(3) to make the final selection of performance indicators to be used in the final model;

(4) to develop and describe procedures (i.e., sources of evidence and membership of the evaluation team) for assessing performance indicators; and,

(5) to make the final selection of incentives to be awarded superior principals in the model.

From on-going conversations at the first and second conferences, two

themes emerged as dominant concerns of the stakeholders that would need to be addressed in constructing the final model. One, the incentives for principals to participate in the evaluation process had to be sufficiently attractive so that principals would be willing to submit to what would be a long and time-consuming evaluation. As one principal put it, "having a plaque in your office looks nice, but what I want is more say about how the money's spent in my school." The question of more local autonomy was a sensitive issue, the very sensitivity of which made it imperative that several administrators from central office be present to discuss this issue on the floor and engage in constructive dialogue with the principals present. The resolution of these discussions was that the incentives for the principals (as voted upon by the entire MDC) consisted of the following: (1) a salary bonus in the amount of $3,500 to be awarded as a lump sum to superior principals at the conclusion of the evaluation (which approximated an 8-12% salary increase based on current pay scales); (2) discretionary funds for educational and school-related purchases which would be tied to the school's enrollment in terms of $10 per child, with the amount to be neither less than $2,00 nor more than $5,000; (3) increased input into the district budgeting process in the form of membership on the district's Budget Development Committee, a group normally comprised of central office staff and school board members; (4) increased autonomy to develop and implement new programs in their respective schools on a pilot basis exclusive of district level approval processes; and, (5) additional public recognition through award ceremonies, media attention and formal recognition at board meetings. This last incentive is the one usually given prominence in most evaluation models, while the preceding four are more meaningful in recognizing at least two components of the superior principal: effective instructional leadership and responsible fiscal management. Not surprisingly, these incentives received the enthusiastic endorsement of the principals on the MDC, while the central office staff, and board members, although initially reluctant to share some of their power, recognized through debates and discussions that such incentives were necessary to maintain high quality, educational leadership by superior principals.

The second theme which shaped the development of the evaluative components of the model was that most of the stakeholders were not technically sophisticated, and furthermore, they did not see the rationale at the local district level for implementing a quantitative model divorced from the behaviors principals actually exhibited in real-life contexts. At the same time, in order for a model to be effective for state-wide implementation, some aspects of the model would have to be quantifiable so that principals' performance across districts could be assessed. The challenge facing the MDC was to construct a model which met both concerns; completing this task was the primary focus of the second and third conferences.

In creating the evaluation component of the model, the first task was the selection of criteria and performance indicators that characterize the superior principal. The process began by expanding upon the items included in a sample instrument developed by the South Carolina State Department of Education, which included 11 criterion categories (leadership, student achievement and development, interpersonal competence, school-community

relations, school climate, personal/professional development, local options, achievement gains, leadership in curriculum development, and staff supervision). Each of these criterion categories contained a set of indicators reflecting aspects of the principal's performance for that criterion. For example, under Leadership, one indicator was: The principal involves the faculty and staff in planning school programs. The MDC was presented with a questionnaire of 75 indicators for all 11 categories and during the first meeting they added 26 more, for a total of 101. They were then asked to rank these indicators on a five point scale as to the appropriateness for identifying superior principals. The analysis of these ratings was the basis for a questionnaire containing 86 indicators that was distributed to 300 persons in the district's educational community and the general community.

After the results of the survey were tabulated, the list of 101 indicators was reduced by eliminating ones that were redundant, or not strongly supported by the MDC and the school and community (as reflected in the mean score assigned to each indicator), and combining indicators that were logically similar in intent. This reduced the listing to 39 indicators, which were presented to the MDC at the second conference. After extensive discussion, revision, and voting, these 39 indicators were reduced to their final form of 25 indicators under 11 criterion categories and served as the basis for the evaluation component of the model.

Once the criterion categories were identified the next step was to identify the types of evidence that would be most appropriate for judging principals' performance in these areas. Four sources of evidence were deemed most appropriate by the MDC: (1) scale instruments, particularly the School Effectiveness Questionnaires developed by the district, which are given to faculty/administrators at the elementary, middle and secondary level, parents, and at the secondary level, to students; (2) on-site observations in the form of a two day visit by an outside evaluation team, who would use a rating checklist to record selected behaviors; (3) interviews conducted by the evaluation team with randomly selected faculty and students in the school, as well as interviews with parents, community leaders, school board members and central office staff; and, (4) documentation in the form of memos to faculty, parental letters, school newsletters, principal's notebook, school records (attendance, resource files, etc.), and personal documents (course transcripts, conference registration forms, etc).

Once consensus from the whole group had been reached as to the incentives principals would receive, and the nature of the evidence that would be collected to assess superior principals, a subset of the MDC, known as the Project Steering Committee, met separately to make final decisions regarding these and other elements (e.g., handling grievances) of the model. One of the committee's major tasks was to consider issues of scaling and weighting with respect to performance indicators included in the model. For scoring purposes, the 11 criterion categories with their accompanying indicators were reorganized into five superordinate dimensions: (1) Student Achievement; (2) Leadership; (3) Staff Supervision; (4) Community Relations; and, (5) Personal Development (see Figure 1). For each indicator within each dimension, two types of evidence would each be ranked on a 5-point scale,

the values ranging from 0-5. In cases where there were more than two types of evidence to be considered, the evaluation team would review all the evidence and select the two best scores in the principal's favor.

Scores from the various sources would have been made comparable by conversion to a common scale as follows:

(a) School Effectiveness Questionnaires and Community Questionnaires - the scores from each measure would be ranked in terms of a percentage of the total score. For example, if the maximum questionnaire score is 200, any score above 80% (160 points) would be given a rank of 5. A score falling below the 60-80 percent range would be given a rank of 4, and so on.

(b) On-site observations - The scores from the rating checklist would be ranked following the same procedures for ranking scores from the questionnaires. If the maximum score from the checklist is 20, then a score of 16 or above would be given a rank of 5, and so on.

(c) Interviews - The evaluation team would use the Delphi technique to reach consensus on how the responses would be ranked, using the same scale from 0-5.

(d) Documentation - The evaluation team would review and rank documents on the same scale as the interview data.

To illustrate to the rest of the MDC how the evaluation would be conducted, case study data was assembled for a fictional principal, Ms. Emmy Lou Harris, principal of Roadville Elementary in a rural South Carolina town. For scoring purposes, weightings had to be assigned to the five performance dimensions and an overall performance criterion established. The weightings, which would normally be established by representatives of the school district using this model, were designated within an overall 240-point maximum score (the maximum number of points for each dimension was obtained by multiplying the number of indicators under each dimension by 10 points) as follows:

(1) Student Achievement	40 point maximum
(2) Leadership	90 point maximum
(3) Staff Supervision	40 point maximum
(4) Community Relations	30 point maximum
(5) Personal Development	40 point maximum

The case study data of Ms. Harris proved a very effective tool for illustrating just how the model would work in an actual setting. The MDC were able to see very clearly just how a rural principal would be able to compete with a colleague from a larger, more urban district by demonstrating how she could amass evidence which drew upon her strengths, ones which may have been weaknesses for principals in larger districts. Since over two-thirds of South Carolina schools were located in rural districts, this issue was particularly important to give all principals a fair chance to win the award.

Another issue the MDC struggled with was the establishment of a cut-off score. The resolution was that during the first year of implementation, the criterion for a superior principal be set at the 75th percentile; that is, obtaining a score of 180 or better (75% of 240) would qualify a principal to receive an incentive award. The rationale for this figure came the State Department of Education, which estimated that as many as 25% of all principals might qualify for a superior rating. Whether this figure would be the most useful benchmark and would need to be adjusted upwards or downward as needed to meet state and local standards was an issue which would have been addressed during the evaluation of the model's first year in operation.

A very critical question for the MDC was the validity of the model in terms of its meaningfulness to the stakeholders. Many participants at these three conferences strongly believed that if an incentive model did not offer real and intangible incentives for principals to compete, and was not grounded in the perceptions and knowledge base of those who were either principals themselves, or in close contact with them (faculty, students, parents, board members), then the model would fail to be useful in any significant way, and thus would not be utilized except under duress by decree from the State Department of Education. The shared feeling at the close of the third conference was that this model was valid for assessing principals, and to further ensure its validity, the notion of a 'superior' principal would be treated as a construct which would have been validated by content analysis procedures of case study data assembled and reviewed by a validation team comprised of state department officials, principals teachers, parents, community leaders, and student representatives. If all of the above named persons are affected by the quality of schooling in their district, then all should have a stake in determining just how a superior principal should be judged.

III. The Evaluator's Role and Training

The making of this model was clearly a collaborative effort on the part of all the MDC members; one could say its elements were jointly constructed and negotiated over the span of three conferences within two months' time. Although this model cannot be characterized in terms of a single model well known in the evaluation literature (e.g., the CIPP model, the Scriven model, etc), the influences both of the stakeholder model and the first author's ethnographic training undoubtedly played a strong role. What was most important is that the evaluators did not play the traditional role of defining goals and objectives for the participants, but instead became, as Kirkup described it, "resources and facilitators whose job it was to develop the skills and confidence of all collective members of the project and to provide whatever support services are necessary for them to achieve what they want" (1986, p. 76). While this role is more easily filled when the evaluation task is formative rather than summative in nature, it is also a forerunner of what Patton (1987) called for in defining the evaluator's role in the future: that of being involved in the "front-end" of project development rather than only being involved in the "back-end" of project evaluation.

For evaluators to be successful in this role, their training must be diversified beyond the bounds of traditional methods of research and

evaluation, which are largely quantitative in orientation. Certainly a command of a broad array of techniques is needed, both qualitative and quantitative, but method alone is not sufficient. Nor is it enough to know and apply a wide range of theories and/or models, although this too would be a prerequisite. What is also required is an understanding of the concept of evaluation in multiple contexts, recognizing the fact that each evaluation will need to be approached differently, and that the evaluator must be skilled in detecting what works in a given situation and be able to supply it (or at least provide coverage through an evaluation team). As Weiss (1986a) perceptively noted,

> "the stakeholder approach changes the role of evaluators. They are not only asked to be technical experts who do competent research. They are required to be political managers who orchestrate the involvement of diverse interest groups. They must be negotiators, weighing one set of information requests against others and coming to amicable agreements about priorities. They must be skillful educators, sharing their knowledge about appropriate expectations for program development and program success while while giving particpants a sense of ownership of the study. Are the expectations for evaluators unreasonably high? (p. 153)

Our conclusion is that the expectations can be met if the preparation of evaluators is modified to meet them. In this sense, the evaluator is not unlike the expert problem solver, who possess a broad array of strategies which can be generalized across problems of different types. Given that our educational system finds it difficult to produce students who can think across domains, we may ask too much of graduate programs in research and evaluation to accomplish this goal, but failure to change will continue to produce evaluators who produce technically perfect reports that sit on the office shelves gathering dust. As one of the teachers commented at the close of the third conference, "I never really understood numbers, but this model makes perfect sense to me in how we're going to evaluate our principals." When models make "perfect sense" to stakeholders, then they will be used, and not before.

References

Kirkup, G. (1986). The feminist evaluator. In E.R. House (Ed.), New directions in educational evaluation (pp. 68-84). London: Falmer Press.

Patton. M.Q. (1986). Utilization-focused evaluations. Beverly Hills, CA: Sage Publications.

Patton, M.Q. (1987). Personal communication. Florida State University, Tallahassee, FL.

Weiss, C. (1986a). The stakeholder approach to evaluation. in E.R. House (Ed.), New directions in educational evaluation (pp. 143-157). Beverly Hills, Ca: Sage Publications.

Weiss, C. (1986b). Toward the future of stakeholder evaluations in the future. In E.R. House (Ed.), New directions in educational evaluation. (pp. 186-198). Beverly Hills, CA: Sage Publications.

Figure 1

The making of a model: Utilizing consensus in formative evaluation

Emihovich, C., Rowls, M., & Lackley, G. Paper presented at the American Evaluation Association Annual Meeting, October 15, Boston.

Dimension 1: Student Achievement

Criterion 1: Achievement Gains

Indicators --- 1. Ensures that students meet basic skills achievement gain/maintenance standards
2. Encourages and supports development of achievement standards in higher order thinking skills

Criterion 1: Student Achievement and Development

Indicator --- 1. Recognizes and rewards effectively individual and group accomplishments of students

Criterion 3: School Climate

Indicators --- 1. Manages appropriately student behavior
2. Ensures that the school plant is an inviting learning environment

Dimension 2: Leadership

Criterion 1: Leadership

Indicators --- 1. Communicates clearly and accurately the school's goals to faculty, staff, students, and parents
2. Involves the faculty, staff and, as appropriate, students in planning school programs
3. Encourages and implements recommend positive changes

Criterion 2: Leadership in Curriculum Development

Indicators --- 1. Ensures that the academic goals of the school are translated into curricular and course objectives
2. Ensures that curricular objective are translated into instructional activities
3. Articulates the school curriculum across grade levels and special programs
4. Ensures that achievement test data are used for the improvement of instruction

5. Encourages and supports development of extracurricular organizations and activities in areas such as music, drama, science fair
6. Encourages and supports the development of students' higher order thinking skills across curricular areas

Dimension 3: Staff Supervision

Criterion 1: Staff Selection, Evaluation and Development

Indicator --- 1. Recognizes and rewards effectively individual and group accomplishments of teachers

Criterion 2: Staff Supervision

Indicators --- 1. Communicates high performance expectations to school staff
2. Assesses effectively performance of school staff
3. Provides for the improvement of staff performance through appropriate staff selection and termination procedures, and staff development procedures

Dimension 4: Community Relations

Criterion 1: School-Community Relations

Indicators --- 1. Implements plans that insure community involvement in and awareness of school programs
2. Presents self well as the chief representative of the school

Criterion 2: Local Options

Indicator --- 1. Recommends policy changes for the improvement of the administrative operation of the school and/or the district

Dimension 5: Personal Development

Criterion 1: Interpersonal Competence

Indicators --- 1. Deals tactfully and fairly with others
2. Manages conflicts effectively
3. Relates effectively with students

Criterion 2: Personal/Professional Development

Indicator --- 1. Keeps abreast of trends, developments and research pertinent to education and school operation

Evaluations proceed as follows:

Step 1: Principals meet eligibility or gateway requirements for participation in the incentive program.

Step 2: Data are collected on principals who meet eligibility requirements on each of the 25 indicators.

Step 3: The data are compiled and scores are totaled into:

1) an overall principal score,
2) a score for each of the 5 dimensions, and
3) a score for each of the 11 criterion categories.

Step 4: Principals who meet the overall criterion score receive the specified awards.

Step 5: Strengths and weaknesses of participating principals are revealed to them based on their high and low dimension and criterion category scores.

Step 6: Provision is made to strengthen principals' weaknesses.